

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

File 351:Derwent WPI 1963-2004/UD,UM &UP=200419

(c) 2004 Thomson Derwent

*File 351: For more current information, include File 331 in your search.
Enter HELP NEWS 331 for details.

Set Items Description

?s pn=jp 10040251 or pn=jp 11342529

1 PN=JP 10040251

1 PN=JP 11342529

S1 2 PN=JP 10040251 OR PN=JP 11342529

?t 1/5/all

1/5/1

DIALOG(R)File 351:Derwent WPI

(c) 2004 Thomson Derwent. All rts. reserv.

012930075 **Image available**

WPI Acc No: 2000-101922/ 200009

XRAM Acc No: C00-030373

XRPX Acc No: N00-078793

Synthetic resin made hollow moulded article - used as container, heat
insulating material, air insulation material for bottles and tanks

Patent Assignee: SEKISUI CHEM IND CO LTD (SEKI)

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
-----------	------	------	-------------	------	------	------

JP 11342529	A	19991214	JP 98359124	A	19981217	200009 B
-------------	---	----------	-------------	---	----------	----------

Priority Applications (No Type Date): JP 9886511 A 19980331

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
-----------	------	-----	----	----------	--------------

JP 11342529	A	8	B29C-049/04		
-------------	---	---	-------------	--	--

Abstract (Basic): JP 11342529 A

NOVELTY - A hollow moulded article (1) of a thermoplastic resin composition contains 1 - 40 wt.% of liquid crystal resin. The liquid crystal resin is distributed in the thermoplastic resin in a fibril state and the fibrils are oriented in a circumferential direction (A) along the outer peripheral surface (2) of the hollow moulded article. The manufacturing method comprises a step for forming a parison by extruding a molten mixture of the liquid crystal resin and the thermoplastic resin through a resin flow passage while rotating a portion of the resin flow passage about a rotation axis agreeing to the resin flow passage direction.

USE - For manufacturing synthetic resin made hollow moulded articles used as container, heat insulating material, air insulation material etc. for bottles, tanks etc..

ADVANTAGE - A synthetic resin made hollow moulded article with excellent mechanical strength and resistance for deformation caused by heat is obtained. DESCRIPTION OF DRAWING(S) - The drawing shows a perspective illustration of the circumferential direction along the outer peripheral surface of a hollow moulded article.

(A)Circumferential direction; (2)Outer peripheral surface.

Dwg.1/4

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-345249

(43)公開日 平成11年(1999)12月14日

(51) Int.Cl.⁸

G O 6 F 17/30

識別記号

FI

G O B F 15/403
15/40

330C
370A

審査請求 未請求 請求項の数22 OL (全 11 頁)

(21)出願番号 特願平11-139845

(22) 出願日 平成11年(1999)5月20日

(31)優先權主張番号 9810795.6

(32)優先日 1998年5月20日

(33)優先権主張国 イギリス (GB)

(71)出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(72)発明者 ピクター ポズナンスキー

イギリス国 オーエックス4 ワイビー

オックスフォード, サンドフォード オ

ン テムズ、 チャーチ ロード 108

(72)発明者 ヤン ヤープ アイデンス

イギリス国 オーエックス1 2オーエフ

オックスフォード、 グローセスター

グリーン、ザ チルターンズ 25

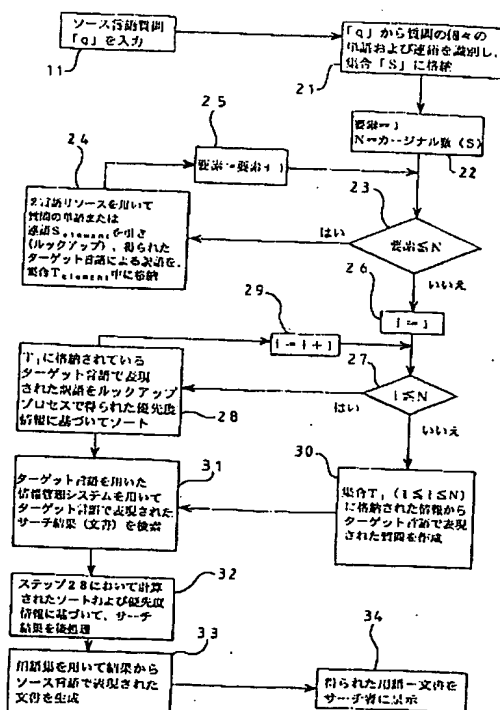
(74)代理人 弁理士 山本 秀策

(54) 【発明の名称】 情報検索のための方法および装置ならびに記憶媒体

(57) 【要約】

【課題】 ターゲット言語を知らないサーチ者であっても検索された文書が関連性を有するものであるか否かをチェックすることができる、言語間情報検索を可能にすること。

【解決手段】 ターゲット言語で書かれた複数の文書から、ソース言語で表現された質問（１１）を用いて情報を検索するための方法が提供される。質問（１１）は、多言語リソース（２１～２９）を用いてターゲット言語に変換される。ソース言語で表現された質問の語の訳語からターゲット言語で表現された質問が形成され（３０）、情報管理システムに適用される（３１）。質問と情報管理システム内の文書の内容との間に一致が発見されると、それぞれの文書またはその一部を、同じ多言語リソースを用いてソース言語に再び変換する。



【特許請求の範囲】

【請求項 1】 ターゲット言語で書かれた複数の文書からソース言語で表現された質問を用いて情報を検索するための方法であって、
多言語リソースを用いて該質問を該ターゲット言語に変換する工程と、
該変換された質問から該ターゲット言語で表現された質問を形成する工程と、
該ターゲット言語で表現された該質問を情報管理システムに適用する工程と、
該情報管理システムによって同定された該ターゲット言語で書かれた文書の少なくとも一部を、該多言語リソースを用いて該ソース言語に変換する工程と、
を包含する、方法。

【請求項 2】 前記ソース言語および前記ターゲット言語は自然言語である、請求項 1 に記載の方法。

【請求項 3】 前記多言語リソースは 2 言語用語集である、請求項 1 または 2 に記載の方法。

【請求項 4】 前記用語集は前記ソース言語で表現された前記質問の各語を識別して翻訳する、請求項 3 に記載の方法。

【請求項 5】 前記用語集は、連語を構成する語を識別して翻訳するが該連語の個々の単語は翻訳しない、請求項 4 に記載の方法。

【請求項 6】 1 つ以上の訳語を有する各語について、前記用語集は該訳語のうち 1 つ以上を供給する、請求項 4 または 5 に記載の方法。

【請求項 7】 前記ターゲット言語で表現された前記質問は、前記ソース言語で表現された質問中の任意の語のうち、前記多言語リソースによって該ターゲット言語に変換され得ないいくつかの語を少なくとも含んでいる、前記請求項のいずれかに記載の方法。

【請求項 8】 各文書の前記少なくとも一部は該文書のタイトルを含む、前記請求項のいずれかに記載の方法。

【請求項 9】 各文書の前記少なくとも一部は該文書の要約または抄録を含む、前記請求項のいずれかに記載の方法。

【請求項 10】 各文書の前記少なくとも一部は、前記ターゲット言語で表現された前記質問に一致する語を含んだ文を含む、前記請求項のいずれかに記載の方法。

【請求項 11】 ターゲット言語で書かれた複数の文書からソース言語で表現された質問を用いて情報を検索するための装置であって、
該質問を該ターゲット言語に変換するための多言語リソースと、
該変換された質問から該ターゲット言語で表現された質問を形成するための手段と、
該ターゲット言語で表現された該質問を情報管理システムに適用する手段とを有しており、
該多言語リソースは、該情報管理システムによって同定

された該ターゲット言語で書かれた文書の少なくとも一部を該ソース言語に変換するように構成された、装置。

【請求項 12】 前記ソース言語および前記ターゲット言語は自然言語である、請求項 11 に記載の装置。

【請求項 13】 前記多言語リソースは 2 言語用語集である、請求項 11 または 12 に記載の装置。

【請求項 14】 前記用語集は前記ソース言語で表現された前記質問の各語を識別して翻訳する、請求項 13 に記載の装置。

【請求項 15】 前記用語集は、連語を構成する語を識別して翻訳するが該連語の個々の単語は翻訳しない、請求項 14 に記載の装置。

【請求項 16】 前記用語集は、1 つ以上の訳語を有する各語について、該訳語のうち 1 つ以上を供給するように構成された、請求項 14 または 15 に記載の装置。

【請求項 17】 前記質問形成手段は、前記ターゲット言語で表現された前記質問中に、前記ソース言語で表現された質問中の任意の語のうち、前記多言語リソースによって該ターゲット言語に変換され得ないいくつかの語を少なくとも含めるように構成された、請求項 11 ないし 16 のいずれかに記載の装置。

【請求項 18】 各文書の前記少なくとも一部は該文書のタイトルを含む、請求項 11 ないし 17 のいずれかに記載の装置。

【請求項 19】 各文書の前記少なくとも一部は該文書の要約または抄録を含む、請求項 11 ないし 18 のいずれかに記載の装置。

【請求項 20】 各文書の前記少なくとも一部は、前記ターゲット言語で表現された前記質問に一致する語を含んだ文を含む、請求項 11 ないし 19 のいずれかに記載の装置。

【請求項 21】 プログラムデータプロセッサを有する、請求項 11 ないし 20 のいずれかに記載の装置。

【請求項 22】 請求項 21 に記載の装置のデータプロセッサを制御するためのプログラムを含んだ、記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報を検索するための方法および装置に関する。本発明はまた、そのような方法を実施するためのプログラムを含んだ記憶媒体に関する。これらの技術は、情報検索システム（「サーチエンジン」）、情報フィルタリングアプリケーション（情報ルーティングシステムとしても知られる）、および情報抽出アプリケーションなどの、情報管理システムにおいて用いられ得る。

【0002】

【従来の技術】D. A. Hull および G. Greffenstette, 「Querying across Languages: a Dictionary-Based Approach

10

20

30

40

50

ach to Multilingual Information Retrieval」、第19回 Annual International Conference on Research and Development in Information Retrieval (SIGIR '96)、第49～57頁、1996に、言語間(cross-linguistic)検索に対する辞書的アプローチの開示がある。選択されたトピックについて関係のある情報を含んだ文書をサーチするために、サーチを行う者(サーチ者)は質問(query)を作製する。典型的な質問は、探したい主題を表す短いテキスト要素、例えば文などを含む。そして、質問が含む単語のうち少なくともいくつかと、各文書のテキスト全体との間の一致を探して、同じ言語で書かれた多数の文書をサーチする。

【0003】この辞書的アプローチにおいて、質問の「ソース」言語とは異なる「ターゲット」言語で書かれた文書をサーチするために、質問の語を2言語辞書を用いてルックアップを行う。ソース言語で表現された質問の各語について可能な訳語の全てを用いてターゲット言語で表現された質問を作成した後、ターゲット言語でマッチング処理を行う。

【0004】従ってこの技術は、サーチ者が、サーチしたい文書の言語とは異なる言語で質問を作成することを可能にする。

【0005】

【発明が解決しようとする課題】質問翻訳技術を用いた公知の言語間検索システムにおいて、質問の全ての語(単語(word)および連語(collocation))をターゲット言語に翻訳し、質問の各語の全ての可能な訳語を用いるか、あるいは、各語の予め決められた(deterministic)好適な訳語を用いて、ターゲット言語で表現された質問を作成する。しかし、これらのいずれのアプローチも以下のような欠点を有している。

【0006】ソース言語で表現された質問の語の全ての可能な訳語を選択することは、その質問に関係しない多くの文書が検索されてしまう結果になり得る。これは、ソース言語の単語は異なる文脈においては異なる意味を有するため、好適な訳語が異なってくるためである。すなわち、典型的な情報システムにおいては多大な量の文書が利用可能であることから、同定された多数の無関係な文書の中からサーチ者が必要な文書を識別することが困難になり得る。

【0007】質問の各語の好適な訳語のみを用いることにより、文書数が増えてしまう問題は解決されるが、公知の機械翻訳システムの精度には限界があり、しばしば不適切な訳語を好適な訳語として選択してしまうことがある。従って、翻訳システムが誤った訳語を選択した場合、求める主題に関係のある文書を情報検索システムが識別できる可能性は少なくなる。

【0008】文脈情報を用いて翻訳精度を改善する試みとして、ソース言語のテキストを分析することにより、連語(同時に呼応して用いられる単語群)を識別する

ための技術が存在する。ある連語に対する可能な訳語の数はその個々の構成単語の可能な訳語の数よりも少ないため、このような処理は意味のある訳語を選択することを助ける。例えば、「make use of」という連語はターゲット言語において少ない訳語しか有さないのに対し、「make」、「use」および「of」という頻繁に用いられる語にはそれぞれ多くの訳語が当てはまる。

【0009】連語の使用は、生成されるターゲット言語の質問の語の数を制限することを助けるが、ほとんどの公知のシステムは、連続的な連語(すなわち隣接した形で同時に呼応して用いられる単語群)を認識することが可能なだけである。実際には、現実の言語における連語のうちのかなりの数が非連続的である。例えば「make use of」という連語は、自然言語で書かれた文書においては「make good use of」のように、「good」という単語をまたがって非連続的な連語になることがある。

【0010】EP 0 813 160号およびGB 2 314 183号は、連続的な連語および非連続的な連語を識別し翻訳するための用語集(glosser)を開示している。「用語集」は、ソース言語の複数の(順序付けられた)単語(または連語)に、ターゲット言語で表現された訳語のラベル付けをすることを可能にする。

【0011】上記に説明したタイプの公知の構成の別の欠点は、識別された文書が、ターゲット言語の状態でサーチ者に呈示されることである。従って、ターゲット言語を知らないサーチ者でも、ターゲット言語で書かれた関係のあるあるいは興味のある文書を検索し得る一方で、ターゲット言語を知らないサーチ者は、検索された文書の関連性および内容をチェックすることができない。従って、公知の言語間情報検索の技術を用い得たとしても、その情報検索の効用は、そのような技術を必要としないほど十分にターゲット言語を知っているサーチ者によってのみしか確かめられ得ないことになる。

【0012】GB 2 320 773号は、主にインターネット上での使用を意図された自動翻訳技術に関する。この技術は、有用な文書またはファイルを文字列に基づいてサーチし、文字列に基づいて探し出された文書に対して最も適切な翻訳環境(例えば用語集または機械翻訳システムなど)を選択することに基づいている。ここで行われる翻訳は、探し出された文書に対してのみ、文字列に基づいて識別された主題に対して最も適切な翻訳環境によって行われる。

【0013】WO 97/18516号は、元の外観を保ちながらウェブページを翻訳することに特に関連する。HTMLコードの周囲にHTMLコードを保存するための概念バリアを置くことによって、HTML文書を前処理する。次にこれらのバリアの外にある残りのテキストおよびデータを、所望の言語に翻訳する。最後に、バリアを除去することにより、ページの元の形式または外観を保存したままで、全ての関連テキストが所望の言語に翻訳された状態にな

5

る。従来質問は通常のアドレスコードで作成されており、それ自体処理を受けることなく、単に所望の文書にアクセスするために用いられるだけであった。

【0014】WO 97/08604号は、質問と文書とを翻訳することに基づいた情報検索システムを開示している。しかしこの技術は、言語から独立した、各質問ならびにサーチに利用可能な各文書の概念的表現を用いている。従ってこのシステムが機能するためには、すべての文書をまず、文書主題の概念的表現を形成するための「翻訳」手順に供さなければならない。質問も同様に処理され、概念的表現のマッチングによりサーチを行う。

【0015】上記問題に鑑み、本発明は、ターゲット言語を知らないサーチ者であっても検索された文書が関連性を有するものであるか否かをチェックすることができ、言語間情報検索を可能にすることを目的とする。

【0016】

【課題を解決するための手段】本発明の第1の局面によれば、ターゲット言語で書かれた複数の文書からソース言語で表現された質問を用いて情報を検索するための方法であって、多言語リソースを用いて該質問を該ターゲット言語に変換する工程と、該変換された質問から該ターゲット言語で表現された質問を形成する工程と、該ターゲット言語で表現された該質問を情報管理システムに適用する工程と、該情報管理システムによって同定された該ターゲット言語で書かれた文書の少なくとも一部を、該多言語リソースを用いて該ソース言語に変換する工程とを包含する方法が提供され、そのことにより上記目的が達成される。

【0017】多言語リソースとは、ソース言語の語（単語または連語）をターゲット言語の1つ以上の同等な語に変換する能力を有する、任意のシステムである。情報管理システムとは、システムに質問として適用された語を含む文書を同定する能力を有する、任意のシステムである。

【0018】前記ソース言語および前記ターゲット言語は自然言語であってもよい。

【0019】前記多言語リソースは2言語用語集であってもよい。

【0020】前記用語集は前記ソース言語で表現された前記質問の各語を識別して翻訳してもよい。

【0021】前記用語集は、連語を構成する語を識別して翻訳するが該連語の個々の単語は翻訳しなくてもよい。

【0022】1つ以上の訳語を有する各語について、前記用語集は該訳語のうち1つ以上を供給してもよい。

【0023】前記ターゲット言語で表現された前記質問は、前記ソース言語で表現された質問中の任意の語のうち、前記多言語リソースによって該ターゲット言語に変換され得ないいくつかの語を少なくとも含んでいてもよい。

6

【0024】各文書の前記少なくとも一部は該文書のタイトルを含んでもよい。

【0025】各文書の前記少なくとも一部は該文書の要約または抄録を含んでもよい。

【0026】各文書の前記少なくとも一部は、前記ターゲット言語で表現された前記質問に一致する語を含んだ文を含んでもよい。

【0027】本発明の第2の局面によれば、ターゲット言語で書かれた複数の文書からソース言語で表現された質問を用いて情報を検索するための装置であって、該質問を該ターゲット言語に変換するための多言語リソースと、該変換された質問から該ターゲット言語で表現された質問を形成するための手段と、該ターゲット言語で表現された該質問を情報管理システムに適用する手段とを有しており、該多言語リソースは、該情報管理システムによって同定された該ターゲット言語で書かれた文書の少なくとも一部を該ソース言語に変換するように構成された、装置が提供され、そのことにより上記目的が達成される。

【0028】前記ソース言語および前記ターゲット言語は自然言語であってもよい。

【0029】前記多言語リソースは2言語用語集であってもよい。

【0030】前記用語集は前記ソース言語で表現された前記質問の各語を識別して翻訳してもよい。

【0031】前記用語集は、連語を構成する語を識別して翻訳するが該連語の個々の単語は翻訳しなくてもよい。

【0032】前記用語集は、1つ以上の訳語を有する各語について、該訳語のうち1つ以上を供給するように構成されてもよい。

【0033】前記質問形成手段は、前記ターゲット言語で表現された前記質問中に、前記ソース言語で表現された質問中の任意の語のうち、前記多言語リソースによって該ターゲット言語に変換され得ないいくつかの語を少なくとも含めるように構成されてもよい。

【0034】各文書の前記少なくとも一部は該文書のタイトルを含んでもよい。

【0035】各文書の前記少なくとも一部は該文書の要約または抄録を含んでもよい。

【0036】各文書の前記少なくとも一部は、前記ターゲット言語で表現された前記質問に一致する語を含んだ文を含んでもよい。

【0037】上記装置は、プログラムドデータプロセッサを有してもよい。

【0038】本発明の第3の局面によれば、上記のいずれかの装置のデータプロセッサを制御するためのプログラムを含んだ記憶媒体が提供され、そのことにより上記目的が達成される。

【0039】用語集は好ましくは、EP 0 813 160号およ

びGB 2 314 183号に開示したタイプのものである。これら出願の内容を本明細書において参考のために援用する。

【0040】上記のように、ターゲット言語を知らないサーチ者であっても検索された文書が関連性を有するものであるか否かをチェックすることが可能な、言語間情報検索を行い得る。質問の作成と、(各)文書の少なくとも一部のソース言語への再変換とを、同じ多言語リソースを用いて行う。従って、変換されたソース言語の文書またはその一部の語が、ソース言語の質問に用いられたのと同じまたは同様な語である可能性が高いという利点が得られる。従って、ターゲット言語を知らないサーチ者であっても、同定されたターゲット言語の文書が実際に質問に関連性を有するものであるか否かを高い精度で決定することができる。従って、サーチ者がターゲット言語を知っているかどうかに関わらず、言語間検索の効率が実質的に改善され得る。

【0041】EP 0 813 160号およびGB 2 314 183号に開示されたもののような、非決定的な(non-deterministic)用語集を用いることの利点は、各語に対する好適な訳語とともに、(例えば利用可能な場合には文の文脈情報を用いることにより)様々な別の訳語案も生成されることである。これにより、生成される訳語案の数が大きく制限される。またこれらの訳語案を、各訳語案が正しい訳語である尤度を示す基準条件に照らして、ランク付けし得る。このようにして、ターゲット言語の質問の形成に実際に用いられる訳語の数を、サーチ者の要求に応じて調整し得る。

【0042】

【発明の実施の形態】以下に本発明を図面を参照して説明する。

【0043】図1は、インターネットサーチエンジンなどの情報管理システム1を用いた、情報検索装置を示す。この装置はコンピュータなどのプログラムドデータプロセッサ型であり、入力インターフェース3(キーボードおよびマウスなど)および出力インターフェース(ディスプレイおよびプリンタなど)を設けたプログラムブルデータプロセッサ2を有する。入力インターフェース3は、ソース言語で表現された質問の入力ならびに、装置の動作をその他制御するために用いられ得る。出力インターフェース4は、サーチの結果の表示ならびに、装置の動作に関する情報を提供するために用いられ得る。

【0044】データプロセッサ2は、データ処理中に一時的にデータを格納するための、ランダムアクセスメモリ(RAM)5の形態である「作業メモリ」を有する。例えば装置の電源がオフにされたときに保持されることが必要なデータを記憶するために、不揮発性読み出し/書き込みメモリ6が設けられている。リードオンリーメモリ(ROM)の形態であるプログラムメモリ7は、デ

ータプロセッサ2の動作を制御するためのプログラムを有している。この装置はさらに、ソース言語で表現された質問の語(単語および連語)に対してターゲット言語の訳語をラベル付けする、文書用語集8を有している。用語集8は好ましくは、EP 0 813 160号およびGB 2 314 183号に開示されたタイプのものである。図において用語集8を装置の独立した要素として示しているが、データプロセッサ2やメモリ5~7によって用語集8を実現してもよい。

10 【0045】プログラムメモリ7は、本明細書において説明される様々な動作を実行するためにデータプロセッサ2によって実行される前述のプログラムおよび/または、文書用語集8を有している。そのようなプログラムは、様々な公知のコンピュータ言語のうち任意の言語で記述され得ることが、コンピュータプログラミングの当業者には理解されるであろう。従って、具体的なコード自体に関する詳細は、簡潔さのために省略する。

20 【0046】情報管理システム1は、機械可読性の文書を含んでおり、質問に基づいてそのような文書にアクセスしあるいはサーチを行うように構成されたタイプのものである。詳細には、システム1は本発明の装置から質問を受け取り、その質問と何らかの一致を有する文書をサーチする。そして、各質問の語または語の論理的な組み合わせを含む文書が例えばメモリ5にダウンロードされることにより、これらの文書を本発明の装置が利用することが可能になる。メモリ5は、十分なメモリ容量を提供するために磁気記憶または光学記憶型のディスクドライブを含み得る。

30 【0047】図2は、図1に示す装置によって行われ得る情報検索の方法を示している。11に示すように、例えば入力インターフェース3のキーボード上で入力されることにより、ソース言語(例えば英語)で表現された質問が作成される。ソース言語の質問は、図1に示す文書用語集8に対応する多言語用語集12に適用される。以下に説明する実施例において、多言語用語集12は、英語の語とオランダ語の語との間の変換を行う、EP 0 813 160号およびGB 2 314 183号に開示されたタイプのものである。

40 【0048】13に示すように、用語集12は、ソース言語で表現された質問の語を、ターゲット言語で表現された質問の語に変換する。また用語集12は、ソース言語質問中のある語に対する複数のターゲット言語質問の語のうちのそれぞれが正しいものである尤度または確率を示す、追加的な情報を供給する。

50 【0049】ターゲット言語質問の語および追加的な情報が質問形成器14に供給され、質問形成器14はこれらの語をターゲット言語質問15に変換する。質問形成器14は追加的な情報を用いることにより、例えば、ターゲット言語質問の語のうち正しい語である確率がしきい値未満であるような語を無視する。ターゲット言語質

問 1 5 は、インターネットサーチエンジンなどの情報管理システム 1 6 に直接適用し得る形式である。

【0050】情報管理システム 1 6 は、マッチング処理を行う。詳細には、システム 1 6 は、ターゲット言語で書かれた文書をサーチすることにより、ターゲット言語質問と文書のテキストとの間の一致を探す。一致が見いだされた文書はターゲット言語結果 1 7 としてダウンロードされる。1 8 において、用語集 1 2 によって提供される追加的な情報を用いることにより、これらの結果を処理する。例えばこれらの結果は再ランク付けされる。10 ランク付けされた結果リスト中において、返される結果のうちで質問の語の「最も尤もらしい(most likely)」訳語を含む結果は、より低い尤もらしさを有する訳語案から生成されたターゲット言語質問の語を含む結果よりも、高い位置に置かれる。そのような再ランク付けを例示する具体的な実施例を後述する。

【0051】次に、処理されたターゲット言語文書の形態である処理化ターゲット言語結果は、多言語用語集 1 2 に供給され、多言語用語集 1 2 は、ターゲット言語結果をソース言語結果 1 9 に変換する。そして、ソース言語結果 1 9 は、出力インターフェース 4 において例えば 20 ディスプレイ上で、またはプリンタからのハードコピーとして、ユーザに提供される。多言語用語集 1 2 は、ターゲット言語で書かれた各文書の全体の用語訳を提供してもよく、あるいは、その一部のみ(タイトル、要約あるいは抄録など)の訳語を提供してもよい。

【0052】図 3 は、図 2 に示す方法の一部をより詳細に示している。本方法の用途を、オランダ語で書かれた文書にアクセスするための、英語で表現されたある特定の 30 ソース言語質問について説明する。

【0053】1 1 に示す入力されたソース言語質問 q は、EP 0 813 160 号および GB 2 314 183 号に開示されたタイプの英語／オランダ語用語集に与えられる。本実施例において、情報管理システム 1 6 は、オランダ語によるワールドワイドウェブサーチエンジンである。ソース言語質問は、「treatments for athlete's foot (水虫の治療法)」であるとする。

【0054】ステップ 2 1 において、用語集は質問の個々の語を識別して集合 S に格納する。これらの語は単語、非連続的な連語、または連続的な連語であってもよい。この特定の質問は以下の語の集合に変換される：

behandelingen AND voor AND (voetschimmel OR (atleet AND (voet OR basis)))

質問形成器 1 4 は、ソース言語質問の語のうち「treatments」と「for」とはオランダ語において単一の訳語を有しており、必ず関係する文書中に存在しているはずであることを検出する。従って、質問形成器 1 4 は、これらの語の訳語をブール論理演算子 AND で関係付けた質問を形成する。

【0057】質問形成器 1 4 は、ソース言語質問の残りの語については様々な可能な訳語が存在することを検出 50

treatments

for

athlete

foot

athlete's foot

ステップ 2 2 において、「要素」パラメータの値を 1 に設定し、パラメータ N を集合 S のカーディナル数、すなわち集合 S の要素の数に等しい値(本具体例においては 5)に設定する。ステップ 2 3 において、「要素」が N 以下であるかどうかをテストする。もし N 以下であれば、ステップ 2 4 において、2 言語リソース(英語／オランダ語辞書または辞典)を用いて質問の語 S_{element} のルックアップを行い、ターゲット言語で表現された(各)訳語を、集合 T_{element} 中に格納する。ステップ 2 5 において、「要素」パラメータを 1 だけインクリメントし、質問の全ての語が翻訳されるまで、ステップ 2 3 から 2 5 を繰り返す。

【0055】ステップ 2 6 においてパラメータ i の値を 1 に設定し、ステップ 2 7 において i が N 以下かどうかをテストする。N 以下であれば、ステップ 2 8 において、集合 T_i に格納されているターゲット言語で表現された訳語を、ステップ 2 4 で行われたルックアッププロセス中に得られた優先度情報に基づいて、ソートする。集合 T_i が 1 つの訳語しか含まない場合は、処理は不要である。しかし、集合 T_i が 1 より多い可能な訳語を含む場合には、それらが正しい語である尤度の順番で訳語はソートされる。

【0056】ステップ 2 9 においてパラメータ i が 1 だけインクリメントされ、入力されたソース言語質問の語の全てのターゲット言語による訳語がソートされるまで、ステップ 2 7 ～ 2 9 を繰り返す。例えば、この入力質問例の場合は、以下のような結果になる：

treatments → [behandelingen]

for → [voor]

athlete → [atleet]

foot → [voet, basis]

athlete's foot → [voetschimmel]

ステップ 3 0 において、集合 T_i (1 ≤ i ≤ N) 中におけるソート済みのターゲット言語による訳語を用いて、ターゲット言語で表現された質問を作成する。ターゲット言語質問の一例は以下ようになる：

する。特に、個々の語「athlete」および「foot」の訳語は両方とも関係する文書中に存在するはずであるが、この代替として「athlete's foot」という語の訳語が存在してもよい。さらに形成器は、「foot」という語の可能な訳語が 2 つ存在するため、これらは質問中において訳語案として規定されなければならないことを検出する。質問中において訳語案はブール論理演算子 OR で関係付けられる。

【0058】ステップ31において質問がターゲット言語を用いた情報管理システムに適用され、ターゲット言語を用いた情報管理システムは、ターゲット言語で表現された文書の形で検索結果を検索する。例えばステッ

1. Moderne behandelingen voor voetschimmel
2. Voetschimmel: nieuwebehandelingen voor een oude kwaal
3. Behandelingen voor aandoeningen van deatleet op basis van nieuwe medische vindingen.

次にステップ32において、ステップ28において計算されたソートおよび優先度情報に基づいて、サーチ結果を処理する。この具体例を以下に示す。

【0059】次いでステップ33において、ターゲット言語で表現されたサーチ結果を用語集12に供給し、用

1. Modern treatments for athlete's foot
2. Athlete's foot: new treatments for an old problem
3. Treatments for injuries of athletes based on new medical discoveries

文書の順序は、最も関係があると考えられる文書を最初の文書として呈示することで、優先度情報を反映させている。この具体例において最初の2つの文書はソース言語質問の語の好適な用語訳を含んでいるため、明らかに関係している。ここで、あまり好適でない訳語の数を制限していることにより、関係のある文書を逃してしまう可能性を減少させ、且つ無関係な文書が探し出されてしまう可能性を減少させている。この具体例において、探

「making good use of old clothes (古着の有効利用)」

この場合、「making use of」は「good」をまたいで非連続的な連語を構成している。このような非連続的な連語を検出することにより、質問をターゲット言語に翻訳する際の精度を改善することが可能である。

【0061】この場合ステップ20～28が提供する、ソース言語質問をターゲット言語に翻訳した訳語のソートされた集合は、以下のようになる：

make use of → [gebruiken, ...]

good → [goed, goedaardig, ...]

old → [oud, antiek, ...]

clothes → [kleren, kledingstukken, ...]

これらの集合は、個々の単語「make」、「use」、および「of」の訳語を含んでいない。従って、連語「make use of」の構成語のうち、おそらくは無関係であるよう

1. Het halen van voldoendes op school: zijnvoldoendes echt voldoende? (英語では「Getting pass marks in school: is a pass really good enough?」に対応)
2. Veiligheid op kantoor:passen en beveiligingsbeambten. (英語では「Safety in the office: passes and security guards」に対応)
3. Een pas opent alle deuren. (英語では「One pass opens all doors」に対応)

ここで質問の語訳としては「passen」の尤度が他よりもはるかに高いにも関わらず、最も尤度の低い訳語を含んだ文書が最も高いランク付けになってしまっている。こ

ブ31は、関係する文書のタイトルまたはそのような文書のうちでターゲット言語質問中の語の必要な組み合わせが存在する一節を、返す。このようなサーチの結果の一例は以下ようになる：

語集12はこれらの結果からソース言語で表現された文書またはテキストを生成する。このようにして、前記のターゲット言語で表現されたサーチ結果は用語集12によって翻訳され、以下の結果となる：

し出されてしまった無関係な文書は1つだけである。

【0060】上記具体例において、用語集は、ソース言語質問の個々の単語に加えて、1個の連続的な連語のみを識別することが必要なだけであった。しかし、EP 0 8 13 160号およびGB 2 314 183号に開示された用語集はまた、非連続的な連語を識別することも可能である。そのような非連続的な連語の具体例は、例えば以下の質問に示される：

な多くの意味が排除されている。この結果、不適切なターゲット言語の語をターゲット言語質問に用いてしまうことで無関係な文書が探し出されてしまう尤度が、実質的に減少する。

【0062】ステップ32で行われる処理の例として、ソース言語質問「introducing security passes」を考える。「passes」のオランダ語訳群およびこれらの語が正しい語である確率は以下のようなものである：

passes	→	pas(sen)	(0.7)
		kaart(en)	(0.2)
		voldoende(s)	(0.1)

これらの訳語案（および当然その他の語の訳語案）から生成される質問は、以下のような結果の集合を返す：

これは、この文書が英語の「passes」の訳語を最も多く含んでおり（質問に対して1個ではなく3個の一致を有している）、サーチエンジンがこれを順序付け基準として

用いているためである。追加的な情報が利用可能ならば、これは望ましい動作ではない。従って、追加的な情報を用い、訳語案の尤度に基づいて結果を再ランク付けする。例えば最も単純な場合において、「文書中の語の発生回数」×「語の尤度」のような式を適用し得る。上

1. Veiligheid op kantoor: passen en beveiligingsbeambten.

2. Een pas opent alle deuren.

3. Het halen van voldoendes op school: zijn voldoendes echt voldoende?

このランキングは、訳語案が正しいものである尤度に関する情報に、よりよく従っている。

【0063】実際において、ソース言語質問の語が、多言語リソースによってターゲット言語質問の語に翻訳し得ない場合が多く存在する。よく見られる例は質問が固有名詞を含む場合である。例えば、多言語リソースが「Dagmar Dwehus」などの固有名詞を翻訳する能力を有することは余り考えられない。しかし、質問におけるそのような語は、関係のある文書を検索する際において非常に有用である（例えば、これらの固有名詞を含む文書のみを検索したい場合など）。

【0064】このような状況への対応策として、多言語リソースが訳語を見いだせないような語は、変更無しにターゲット言語質問に渡されてもよい。このような語は以下のようにターゲット言語質問で用いられる。すなわち、1) これらの語を含む文書のみが検索されるか、あるいは2) これらの語を含む文書およびこれらの語を含まないが他のサーチ基準を満たす文書が検索されるように、用いられる。

【0065】別態様として、そのような「翻訳不可能な」語は、ターゲット言語質問の作成中に編集されてもよい。例えば、そのような語は、固有名詞と識別された場合のみターゲット言語質問に導入されるようにしてもよい。その結果、固有名詞と識別されなかった語は捨てられ、ターゲット言語質問の部分構成することがなくなる。この方法は、ソース言語質問で綴りを誤ったために翻訳不可能になってしまった語を捨捨棄するために用い得る。

【0066】ターゲット言語質問を形成する方法論を決定する際において、多言語リソースの実際の形態を考慮することが必要になり得る。例えば、機械翻訳システムは、ターゲット言語質問の語の訳として非常に希な訳を生成する可能性を有するが、一方、用語集はより普遍的な訳語を生成し得る。これは全体的な検索精度に効果を有するものと考えられる。

【0067】例えば、その効果の一つとして、情報検索システムは統計的な性質を有することを利用して、普遍的な語であればあるほど検索システムに挙がってくる文書が少なくなるようにすることで、他の語のために文書が挙がってくる余地を拡大することが可能である。この結果、ターゲット言語質問の語が誤訳であった場合でも、検索結果に対する悪影響は少なくなる。これに対し

記の例においては、1番目の文書の新しい点数は $0.3(3 \times 0.1)$ となり、残りの2つの文書はそれぞれ $0.7(1 \times 0.7)$ となり、従って新しいランキングは以下の通りである：

10 の誤りに起因する悪影響がより大きくなる。なぜなら、結果として得られる語が希であることにより、検索システムにおいてより大きな重要性が与えられてしまい、従って無関係である可能性のある文書がより多く結果に挙がってきてしまい得るからである。

【0068】ターゲット言語質問中において最も普遍的に用いられる語を使用することにより、関係する可能性の高い文書を数多く検索してくることを助け得る。またこのような方法によって関係する文書が多く検索されすぎることはないと考えられる。なぜなら、語が普遍的なものであるということは、情報検索システムにとって、関係ある文書と無関係な文書とを区別する際におけるその価値が恐らく非常に低いであろうというヒントになるからである。

【0069】多言語リソースの性質によっては、ソース言語で表現された各語の「デフォルト訳語」を選択することによりターゲット言語における対応する語を作成することでも十分である。例えば多言語リソースが2言語辞書である場合、このデフォルト訳語は、その辞書から得られる好適な訳語であってもよい。例えば多数の文書を分析することなどにより、ある語の発生頻度に関する情報が得られている場合は、最も普遍的な語をデフォルト訳語として選択し得る。

【0070】

【発明の効果】本発明によれば、ターゲット言語を知らないサーチ者であっても検索された文書が関連性を有するものであるか否かをチェックすることができる、言語間情報検索が可能となる。

【0071】すなわち、本発明によれば、質問の作成と、文書の少なくとも一部のソース言語への再変換とを、同じ多言語リソースを用いて行うことにより、変換されたソース言語の文書またはその一部の語が、ソース言語の質問に用いられたのと同じまたは同様な語である可能性を高くすることができる。これにより、ターゲット言語を知らないサーチ者であっても、同定されたターゲット言語の文書が実際に質問に関連性を有するものであるか否かを高い精度で決定することができる。従って、サーチ者がターゲット言語を知っているかどうかに関わらず、言語間検索の効率が実質的に改善され得る。

【0072】また、本発明によれば、非決定的な用語集を用いることにより、各語に対する好適な訳語とともに

に、様々な別の訳語案も生成される。これにより、生成される訳語案の数が大きく制限され、これらの訳語案を、各訳語案が正しい訳語である尤度を示す基準条件に照らしてランク付けし得る。このようにして、ターゲット言語の質問の形成に実際に用いられる訳語の数を、サーチ者の要求に応じて調整し得る。

【図面の簡単な説明】

【図 1】 図 1 は、本発明の一実施形態における情報検索装置の、概略ブロック図である。

【図 2】 図 2 は、本発明の一実施形態における情報検索方法の、フロー図である。

【図 3】 図 3 は、図 2 に示した方法と同様な方法の、より詳細なフロー図である。

【符号の説明】

1、16 情報管理システム

2 プログラマブルデータプロセッサ

3 入力インターフェース

4 出力インターフェース

5 「作業」メモリ

6 不揮発性読み出し／書き込みメモリ

7 プログラムメモリ

8 文書用語集

11 ソース言語質問

12 複数言語用語集

13 ターゲット質問の語および追加的な情報

14 質問形成器

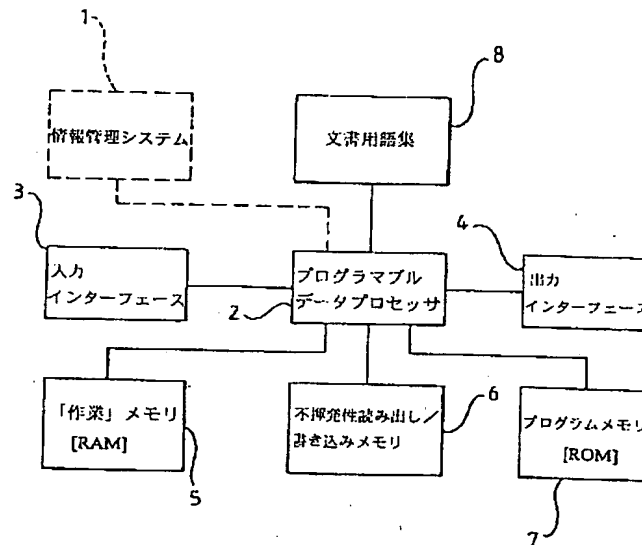
15 ターゲット言語質問

17 ターゲット言語結果

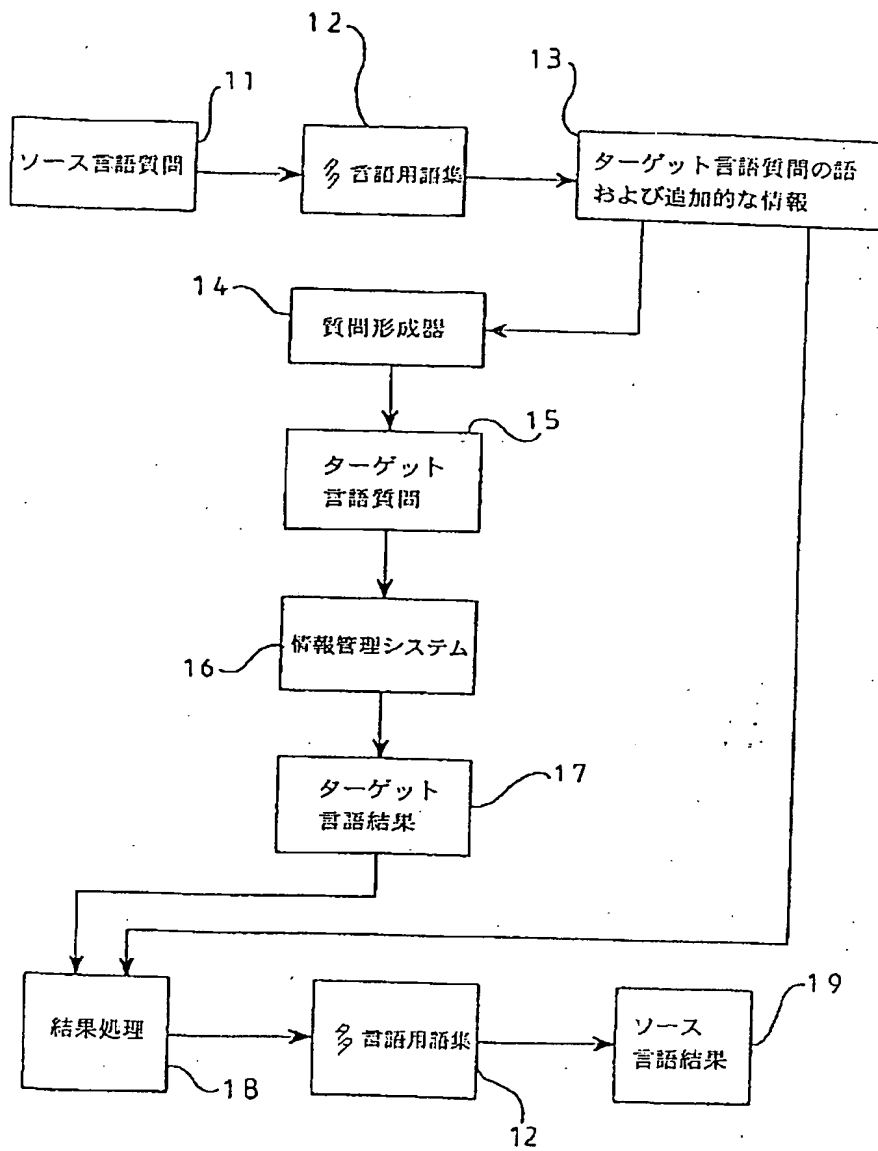
18 結果処理

19 ソース言語結果

【図 1】



【図 2】



【図 3】

